RAW SEQUENCE LISTING
PATENT APPLICATION   *US/09/017,715A*

INPUT SET: S32850.raw

> **This Raw Listing contains the General**
> **Information Section and up to the first 5 pages.**

```
1                                SEQUENCE LISTING
2
3    (1)    General Information:
4
5         (i) APPLICANTS: Ji, Hongjun
6                         Rosen, Craig A.
7
8        (ii) TITLE OF INVENTION: Breast Cancer Specific Gene 1
9
10      (iii) NUMBER OF SEQUENCES: 12
11
12       (iv) CORRESPONDENCE ADDRESS:
13             (A) ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
14             (B) STREET: 1100 New York Ave., Suite 600
15             (C) CITY: Washington
16             (D) STATE: DC
17             (E) COUNTRY: USA
18             (F) ZIP: 20005-3934
19
20        (v) COMPUTER READABLE FORM:
21             (A) MEDIUM TYPE: Floppy disk
22             (B) COMPUTER: IBM PC compatible
23             (C) OPERATING SYSTEM: PC-DOS/MS-DOS
24             (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
25
26       (vi) CURRENT APPLICATION DATA:
27             (A) APPLICATION NUMBER: US 09/017,715
28             (B) FILING DATE: 1998-FEB-03
29             (C) CLASSIFICATION:
30
31      (vii) PRIOR APPLICATION DATA:
32             (A) APPLICATION NUMBER: US 08/673,284
33             (B) FILING DATE: 28-JUN-96
34             (C) CLASSIFICATION:
35
36      (vii) PRIOR APPLICATION DATA:
37             (A) APPLICATION NUMBER: US 60/000,602
38             (B) FILING DATE: 30-JUN-95
39             (C) CLASSIFICATION:
40
41      (vii) PRIOR APPLICATION DATA:
42             (A) APPLICATION NUMBER: PCT/US95/08295
43             (B) FILING DATE: 30-JUN-95
44             (C) CLASSIFICATION:
45
46      (vii) PRIOR APPLICATION DATA:
```

ENTERED

```
47              (A) APPLICATION NUMBER: US 60/037,080
48              (B) FILING DATE: 03-FEB-97
49              (C) CLASSIFICATION:
50
51       (viii) ATTORNEY/AGENT INFORMATION:
52              (A) NAME: Steffe, Eric K.
53              (B) REGISTRATION NUMBER: 36,688
54              (C) REFERENCE/DOCKET NUMBER: 1488.0810003
55
56         (ix) TELECOMMUNICATION INFORMATION:
57              (A) TELEPHONE: 202-371-2600
58              (B) TELEFAX: 202-371-2540
59
60
61   (2) INFORMATION FOR SEQ ID NO:1:
62
63       (i) SEQUENCE CHARACTERISTICS:
64              (A) LENGTH: 550 base pairs
65              (B) TYPE: nucleic acid
66              (C) STRANDEDNESS: double
67              (D) TOPOLOGY: both
68
69      (ii) MOLECULE TYPE: cDNA
70
71
72      (ix) FEATURE:
73              (A) NAME/KEY: CDS
74              (B) LOCATION: 12..392
75
76
77      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
78
79   CACGAGCCAC C ATG GAT GTT TTC AAG AAG GGC TTC TCC ATC GCC AAG AAG          50
80                Met Asp Val Phe Lys Lys Gly Phe Ser Ile Ala Lys Lys
81                 1           5                    10
82
83   GGC GTG GTG GGT GCG GTG GAA AAG ACC AAG CAG GGG GTG ACG GAA GCA          98
84   Gly Val Val Gly Ala Val Glu Lys Thr Lys Gln Gly Val Thr Glu Ala
85      15              20                  25
86
87   GCT GAG AAG ACC AAG GAG GGG GTC ATG TAT GTG GGA GCC AAG ACC AAG         146
88   Ala Glu Lys Thr Lys Glu Gly Val Met Tyr Val Gly Ala Lys Thr Lys
89    30              35                  40                      45
90
91   GAG AAT GTT GTA CAG AGC GTG ACC TCA GTG GCC GAG AAG ACC AAG GAG         194
92   Glu Asn Val Val Gln Ser Val Thr Ser Val Ala Glu Lys Thr Lys Glu
93                  50                  55                  60
94
95   CAG GCC AAC GCC GTG AGC AAG GCT GTG GTG AGC AGC GTC AAC ACT GTG         242
96   Gln Ala Asn Ala Val Ser Lys Ala Val Val Ser Ser Val Asn Thr Val
97                  65                  70                  75
98
99   GCC ACC AAG ACC GTG GAG GAG GCG GAG AAC ATC GCG GTC ACC TCC GGG         290
```

INPUT SET: S32850.raw

```
100   Ala Thr Lys Thr Val Glu Glu Ala Glu Asn Ile Ala Val Thr Ser Gly
101       80                  85                  90
102
103   GTG GTG CGC AAG GAG GAC TTG AGG CCA TCT GCC CCC CAA CAG GAG GGT        338
104   Val Val Arg Lys Glu Asp Leu Arg Pro Ser Ala Pro Gln Gln Glu Gly
105       95                  100                 105
106
107   GAG GCA TCC AAA GAG AAA GAG GAA GTG GCA GAG GAG GCC CAG AGT GGG        386
108   Glu Ala Ser Lys Glu Lys Glu Glu Val Ala Glu Glu Ala Gln Ser Gly
109   110                 115                 120                 125
110
111   GGA GAC TAGAGGGCTA CAGGCCAGCG TGGATGACCT GAAGAGCGCT CCTCTGCCTT        442
112   Gly Asp
113
114
115   GGACACCATC CCCTCCTAGC ACAAGGAGTG CCCGCCTTGA GTGACATGCG GGTGCCCACG      502
116
117   CTCCTGCCCT CGTCTCCCTG GACACCCTTG GCCTGTCCAC CTGTGCTG                   550
118
119
120   (2) INFORMATION FOR SEQ ID NO:2:
121
122          (i) SEQUENCE CHARACTERISTICS:
123                 (A) LENGTH: 127 amino acids
124                 (B) TYPE: amino acid
125                 (D) TOPOLOGY: linear
126
127          (ii) MOLECULE TYPE: protein
128
129          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
130
131   Met Asp Val Phe Lys Lys Gly Phe Ser Ile Ala Lys Lys Gly Val Val
132    1               5                  10                  15
133
134   Gly Ala Val Glu Lys Thr Lys Gln Gly Val Thr Glu Ala Ala Glu Lys
135               20                  25                  30
136
137   Thr Lys Glu Gly Val Met Tyr Val Gly Ala Lys Thr Lys Glu Asn Val
138           35                  40                  45
139
140   Val Gln Ser Val Thr Ser Val Ala Glu Lys Thr Lys Glu Gln Ala Asn
141       50                  55                  60
142
143   Ala Val Ser Lys Ala Val Val Ser Ser Val Asn Thr Val Ala Thr Lys
144    65                  70                  75                  80
145
146   Thr Val Glu Glu Ala Glu Asn Ile Ala Val Thr Ser Gly Val Val Arg
147               85                  90                  95
148
149   Lys Glu Asp Leu Arg Pro Ser Ala Pro Gln Gln Glu Gly Glu Ala Ser
150               100                 105                 110
151
152   Lys Glu Lys Glu Glu Val Ala Glu Glu Ala Gln Ser Gly Gly Asp
```

RAW SEQUENCE LISTING
PATENT APPLICATION   *US/09/017,715A*

INPUT SET: S32850.raw

```
153              115                   120                     125
154
155     (2) INFORMATION FOR SEQ ID NO:3:
156
157          (i) SEQUENCE CHARACTERISTICS:
158              (A) LENGTH: 23 base pairs
159              (B) TYPE: nucleic acid
160              (C) STRANDEDNESS: single
161              (D) TOPOLOGY: linear
162
163         (ii) MOLECULE TYPE: cDNA
164
165
166
167
168         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
169
170     GGGGATCCAT GTTTTCAAGA AGG                                      23
171
172     (2) INFORMATION FOR SEQ ID NO:4:
173
174          (i) SEQUENCE CHARACTERISTICS:
175              (A) LENGTH: 27 base pairs
176              (B) TYPE: nucleic acid
177             .(C) STRANDEDNESS: single
178              (D) TOPOLOGY: linear
179
180         (ii) MOLECULE TYPE: cDNA
181
182
183
184
185         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
186
187     GGAAGCTTCT AGTCTCCCCC ACTCTGG                                  27
188
189     (2) INFORMATION FOR SEQ ID NO:5:
190
191          (i) SEQUENCE CHARACTERISTICS:
192              (A) LENGTH: 25 base pairs
193              (B) TYPE: nucleic acid
194              (C) STRANDEDNESS: single
195              (D) TOPOLOGY: linear
196
197         (ii) MOLECULE TYPE: cDNA
198
199
200
201
202         (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
203
204     GGGGATCCCG ATGTTTTCAA GAAGG                                    25
205
```

INPUT SET: S32850.raw

```
206    (2) INFORMATION FOR SEQ ID NO:6:
207
208        (i) SEQUENCE CHARACTERISTICS:
209            (A) LENGTH: 27 base pairs
210            (B) TYPE: nucleic acid
211            (C) STRANDEDNESS: single
212            (D) TOPOLOGY: linear
213
214        (ii) MOLECULE TYPE: cDNA
215
216
217
218
219        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
220
221    GGGGTACCCT AGTCTCCCCC ACTCTGG                                    27
222
223    (2) INFORMATION FOR SEQ ID NO:7:
224
225        (i) SEQUENCE CHARACTERISTICS:
226            (A) LENGTH: 29 base pairs
227            (B) TYPE: nucleic acid
228            (C) STRANDEDNESS: single
229            (D) TOPOLOGY: linear
230
231        (ii) MOLECULE TYPE: cDNA
232
233
234
235
236        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
237
238    GGGGATCCGC CACCATGTTT TCAAGAAGG                                  29
239
240    (2) INFORMATION FOR SEQ ID NO:8:
241
242        (i) SEQUENCE CHARACTERISTICS:
243            (A) LENGTH: 60 base pairs
244            (B) TYPE: nucleic acid
245            (C) STRANDEDNESS: single
246            (D) TOPOLOGY: linear
247
248        (ii) MOLECULE TYPE: cDNA
249
250
251
252
253        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
254
255    GGGGATCCTC AGAAAGCGTA GTCTGGGACG TCGTATGGGT ACTAGTCTCC CCCACTCTGG   60
256
257
258    (2) INFORMATION FOR SEQ ID NO:9:
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION *US/09/017,715A*

INPUT SET: S32850.raw

Line	Error	Original Text